



Module 1:

Introduction to Generative AI

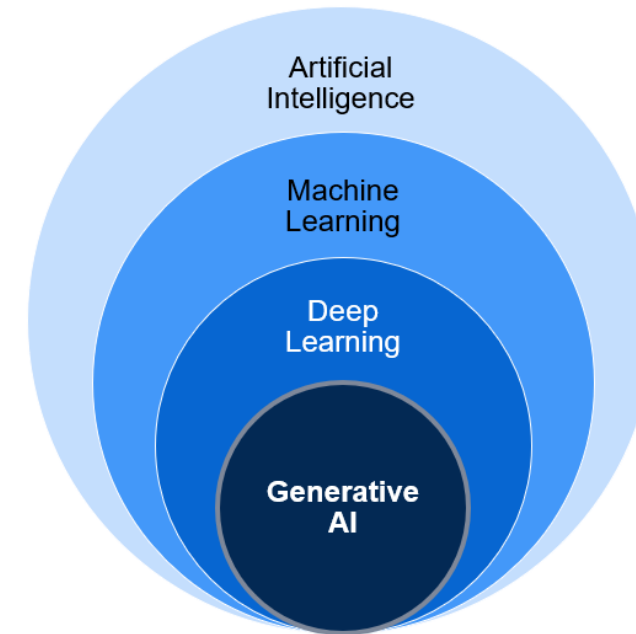
307307 – BI Methods and Models



The AI Landscape

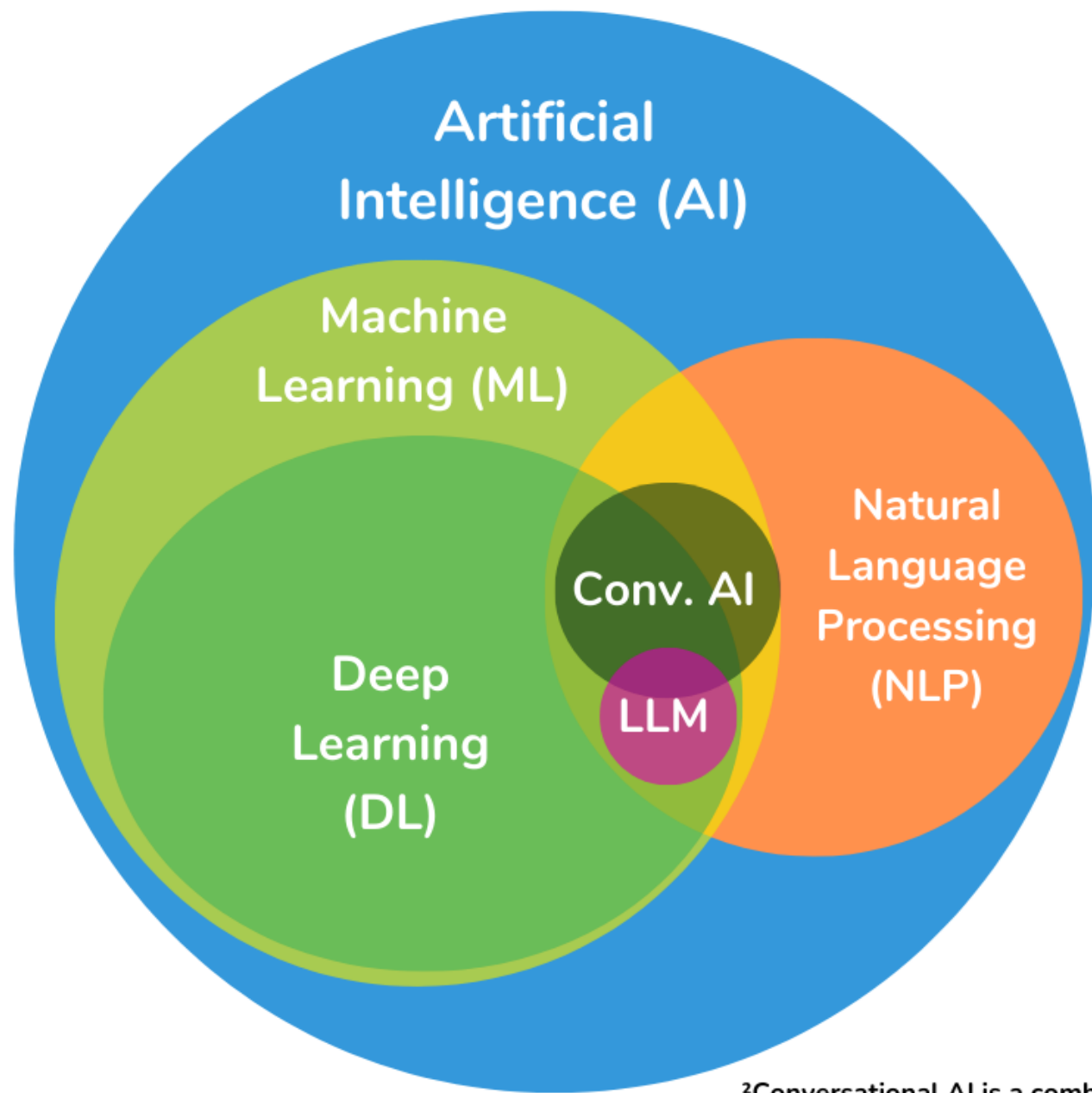
- **Artificial Intelligence (AI)** refers to **any technique that enables machines to mimic human intelligence**. It is the broadest field, including all computational methods that allow machines to reason, learn, and make decisions like humans.
- **Machine Learning (ML)** is a **subset of AI that gives machines the ability to learn from data** without being explicitly programmed. Instead of following hard-coded rules, ML systems improve their performance by recognizing patterns and making predictions based on past observations.
- **Deep Learning (DL)** is a **specialized branch of ML that uses multi-layered neural networks to automatically extract complex patterns from data**. Deep learning has driven major advances in areas such as computer vision, speech recognition, and natural language processing.
- **Large Language Models (LLMs)** are a **type of deep learning model trained on massive amounts of text data**. They are designed to understand, generate, and respond to human language, forming the foundation of Generative AI systems such as ChatGPT.

The AI Landscape



In essence, Generative AI builds upon this hierarchy:
AI → Machine Learning → Deep Learning → Large Language Models.

Each layer narrows in scope but increases in sophistication, culminating in systems capable of producing human-like language and creative outputs.



- Artificial Intelligence (AI)
- Machine Learning (ML)
- Deep Learning (DL)
- Natural Language Processing (NLP)
- Large Language Model (LLM)¹
- Conversational AI (Conv. AI)²

¹LLM is an intersection of DL and NLP

²Conversational AI is a combination of ML and NLP. It may include DL and LLM, but that isn't always the case.

What is Generative AI?

Generative AI is a different class of AI.

Instead of just analyzing data, it's designed to **create new, original content** that has never existed before. This content can take many forms:

- **Text:** Articles, emails, code, poems, scripts, and marketing copy.
- **Images:** Photos, digital art, logos, and product designs.
- **Audio:** Music, voiceovers, and sound effects.
- **Video:** Short clips, animations, and movie scenes.

Why is it called "Generative"?

- The term "generative" simply means "capable of producing or creating."
- The AI doesn't just retrieve information; it synthesizes it in new and imaginative ways to generate something novel.
- This is a massive leap forward, as it moves AI from a tool for analysis to a partner for **creation and innovation**.

Discriminative vs Generative

	Discriminative Model (the judge)	Generative Model (the artist)
Goal	Classify data and distinguish between classes.	Generate new data similar to the originals.
Question	Is this a cat or a dog?	Make me a picture of a cat!
Examples	Spam filters, medical diagnosis, face detection.	ChatGPT, DALL-E, music models.

Why Does Generative AI Matter for Business?

Generative AI is a powerful tool for driving business value. Its ability to create content can:

- **Business Applications:** It enhances customer service through AI-powered **chatbots**, automates report generation, and brings in optimizations in various processes in terms of cost and manual effort.
- **Increase Efficiency:** It can **automate** time-consuming, repetitive **tasks** like drafting emails, writing product descriptions, or summarizing documents, allowing employees to focus on higher-value work.
- **Unlock Creativity and Innovation:** It can act as a brainstorming partner, **generating new ideas** for products, marketing campaigns, or even business strategies. This can accelerate the creative process from days to minutes. Also, generative AI is used in art, music, and writing to create new pieces that mimic human creativity. Tools like DALL-E and ChatGPT are popular examples.
- **Personalize at Scale:** Businesses can use it to **create highly personalized content** for individual customers, such as customized marketing emails or product recommendations, at a massive scale that was previously impossible.
- In essence, generative AI is **a new type of intelligent automation** that can handle complex, creative tasks.
- For a business leader, understanding how to apply this technology is key to gaining **a competitive advantage**.

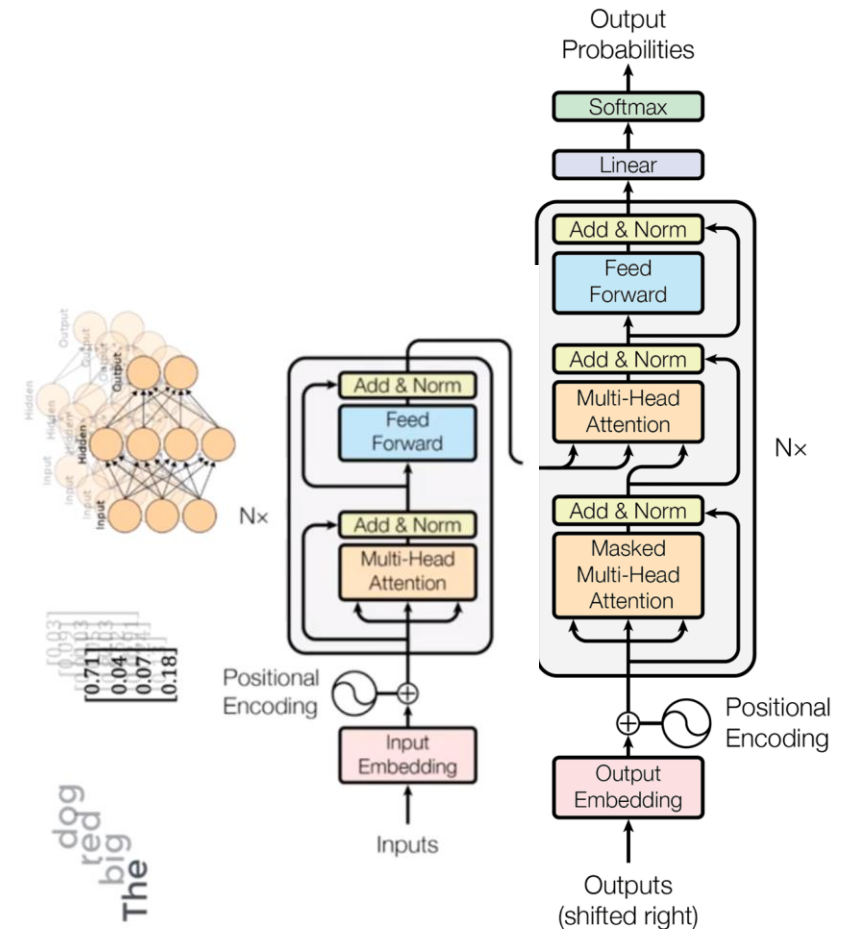
Large Language Models (LLMs)

Large Language Models (LLMs)

- **LLMs** are the **engines** that power popular tools like ChatGPT, Gemini, and Claude.
- Large Language Model is a massive **Neural Networks** that has been trained on an enormous amount of text and data from the internet—books, articles, websites, and more.
- This training process isn't about memorizing facts; it's about learning the **patterns, grammar, and relationships** within language.
- These models can understand and generate human language, enabling tasks like answering questions, summarizing text, or writing content.

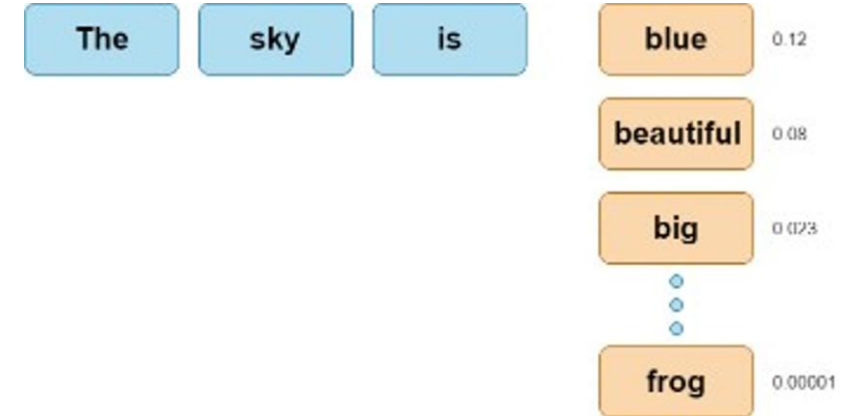
Large Language Models (LLMs)

- Large Language Models are built on the **Transformer Architecture**, which is a neural network architecture that is **well-suited for processing sequences**.
- Transformer architecture is a LARGE **Neural Network** that can reach thousands of neurons and billions of weights.
- These models are trained to learn patterns in human language, enabling them to generate coherent text and support more natural, intuitive interactions with technology.



How do LLMs work?

- LLMs function as complex auto-completion systems.
- They are designed to suggest the most likely next word based on their exposure to similar contexts during training.
- Consider this example:
when given the phrase “The sky is”, an LLM predicts the most likely next word based on learned patterns.
- “Blue” might rank highest, while words like “beautiful” or “big” are also plausible.
- Less likely words, like “frog”, may still receive some probability, even if they don’t fit the context.



- LLMs operate based on probabilistic knowledge, predicting words and phrases according to patterns in data rather than true understanding.
- They do not possess a semantic understanding of the content they generate.

Why are LLMs called large language models?

- LLMs are called "large" because of their immense scale in both architecture and training data.
- LLMs consist of **billions to trillions** of trainable parameters—internal values the model learns and adjusts to identify patterns and generate accurate outputs.
- In general, the larger the model, the better its performance on a wide range of language tasks.



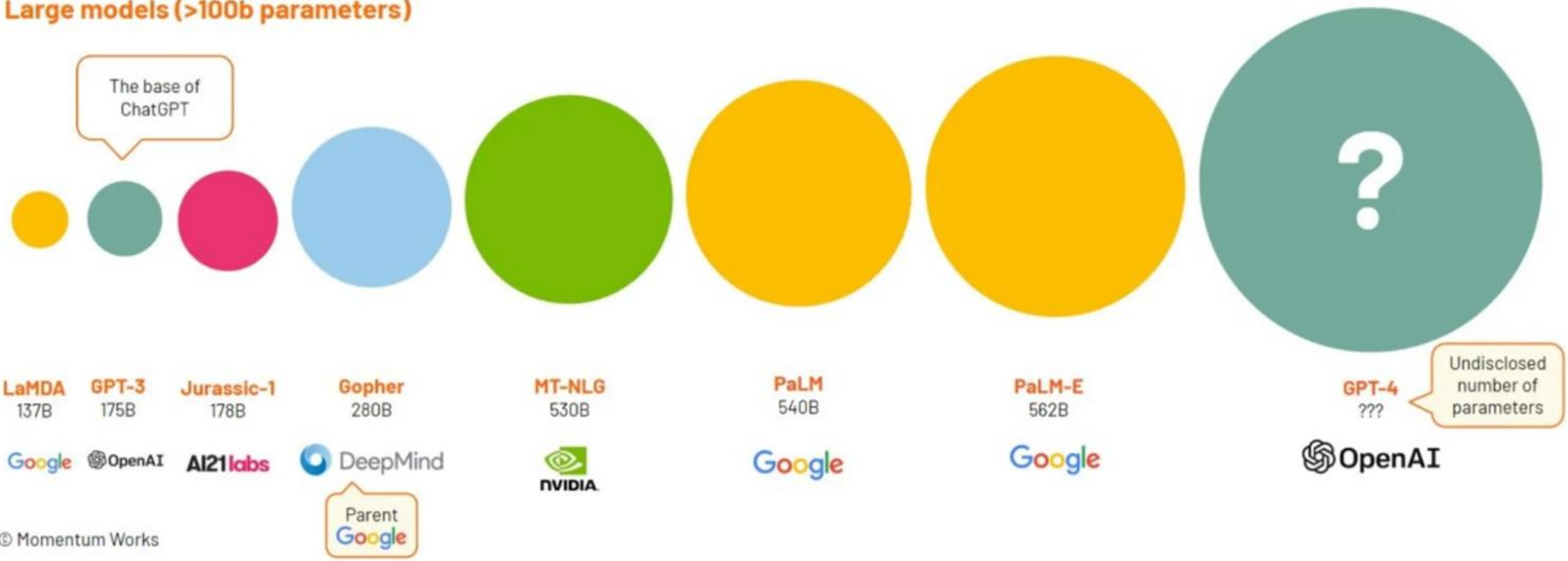
For example, OpenAI's GPT-3 has 175 billion parameters, while GPT-4 is estimated to have around 1.76 trillion, demonstrating how size correlates with capability.

- In addition to their size, LLMs are trained on *massive datasets* and require *substantial computational resources*, making it impractical for most users to train them from scratch.
- This combination of vast parameter counts, massive training data, and advanced computing power is what makes these models truly large.

Small models (<= 100b parameters)



Large models (>100b parameters)



Generative AI

Generative AI Modalities and Their Business Applications

- GenAI models can generate:

1. Text
2. Images
3. Audio
4. Video

Each modality unlocks unique business value:

- **Text Generation:** Chatbots, content creation, summarization, translation, sentiment analysis, report drafting.
- **Image Generation:** Marketing visuals, product design mockups, branding materials, medical imaging insights, and creative concept visualization.
- **Audio Generation:** Voice assistants, personalized audio ads, synthetic narration, multilingual voiceovers, and accessibility tools.
- **Video Generation:** Automated training videos, marketing campaigns, product demos, virtual influencers, and visual storytelling.

Image Generation

- **Definition:** Much like LLMs are the leading type of generative AI models for NLP tasks, diffusion models are the state-of-the-art approach for generating visual content like images and art.
- The principle behind diffusion models is to **gradually add noise to an image and then learn to reverse this process through denoising**.
- By doing so, the model learns highly intricate patterns, ultimately becoming capable of creating impressive images that often appear photorealistic.

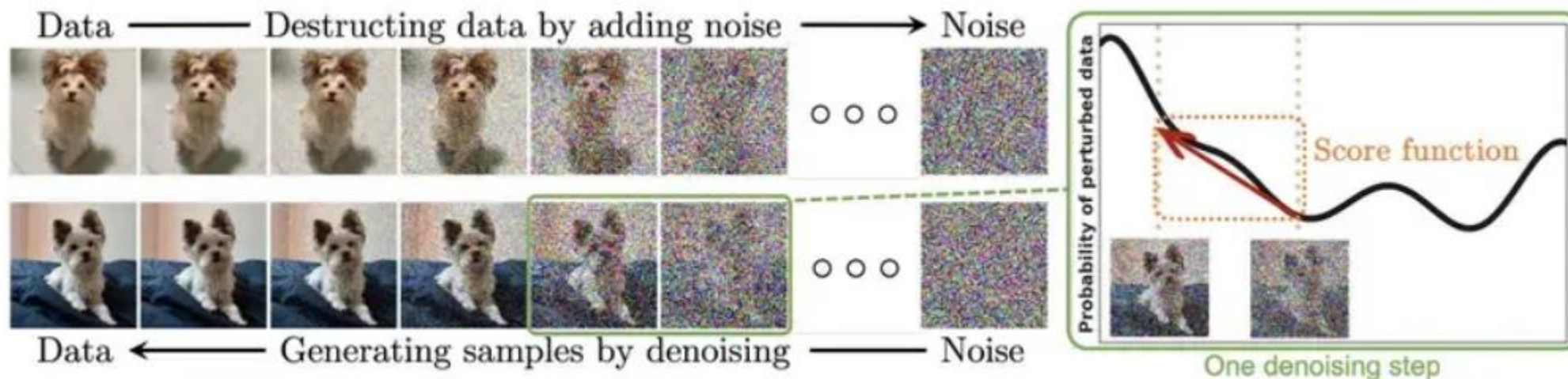


Image Generation

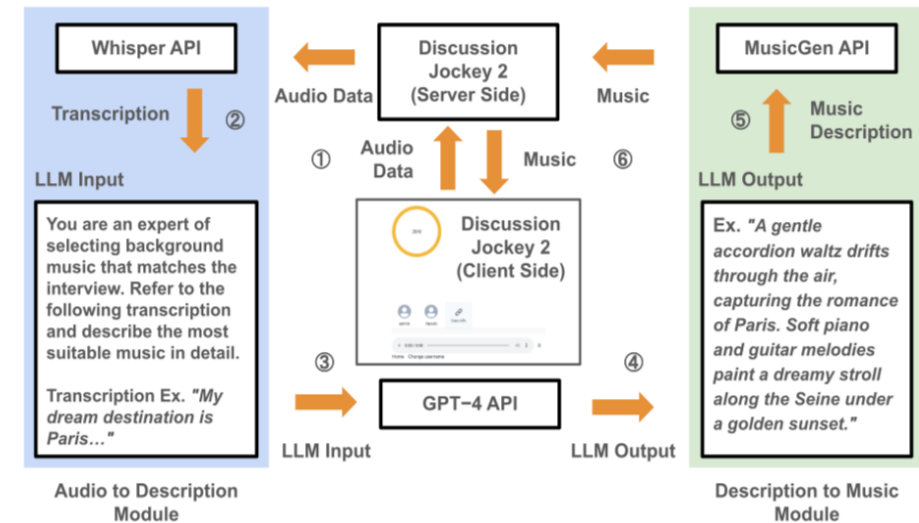
- Models like **DALL-E**, **Midjourney**, and **Stable Diffusion** are trained on a massive dataset of images and their corresponding text descriptions.
- When you give them a prompt like "a photorealistic image of a cat in a business suit drinking coffee," the model generates a new, original image that matches the description.
- **Business Applications:**
 - **Marketing & Advertising:** Quickly create unique visuals for social media campaigns, ad banners, or blog posts without a designer.
 - **Product Design:** Generate new product mockups or explore different design options in a fraction of the time.
 - **E-commerce:** Create lifelike product photos or virtual try-on experiences without expensive photo shoots.



1. <https://openai.com/index/dall-e-2/>
2. <https://www.midjourney.com/home>

Audio Generation (MUSIC and Speech)

- These models can create original music, sound effects, or lifelike human speech.
- They are trained on vast libraries of audio data and can generate new compositions based on a text description or a musical theme.
- **Business Applications:**
 - **Content Creation:** Generate royalty-free background music for a video or a podcast.
 - **Customer Experience:** Create natural-sounding voiceovers for virtual assistants or automated phone systems.
 - **Gaming:** Dynamically create sound effects and soundtracks that adapt to the game's environment.



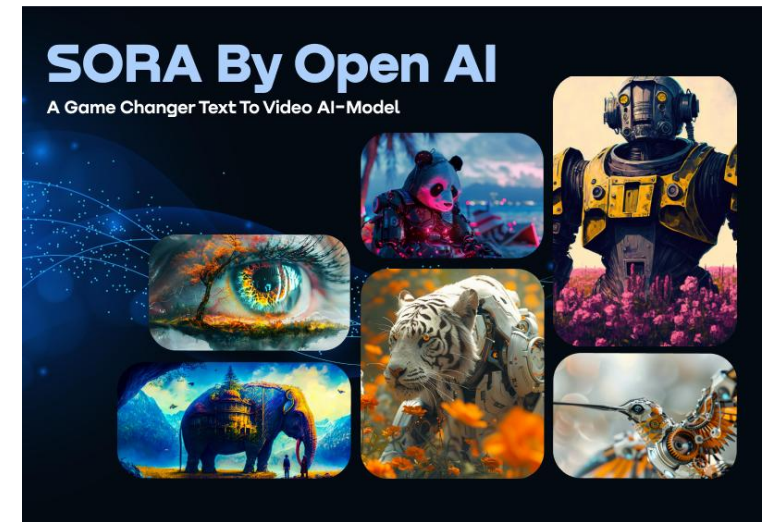
The diagram shows a workflow for automatically generating background music that matches spoken dialogue using AI models.

1. Whisper API transcribes spoken audio into text.
2. The transcription is passed into the GPT-4 API, which acts as an Audio-to-Description Module. It analyzes the context and creates a descriptive explanation of the most suitable music style.
3. This music description (e.g., "a gentle accordion waltz...") is sent to MusicGen API, the Description-to-Music Module, which generates actual music.
4. The Discussion Jockey 2 (Server + Client Side) coordinates the flow of audio data, transcription, music description, and generated music.

Essentially, the system listens to speech, interprets the context, and then produces matching background music automatically.

Video Generation

- Models like **Sora** can be used to create high-quality, realistic video clips from simple text prompts.
- **Business Applications (Emerging):**
 - **Entertainment:** Rapidly prototype animated storyboards for films or create short marketing videos.
 - **Training:** Generate realistic simulations for employee training, such as safety or machinery operation videos.



1. <https://sora.chatgpt.com/>

Google AI Studio Overview



What is Google AI Studio?

- Google AI Studio is a web-based platform for building and testing generative AI applications using Google's Gemini models.
- It supports multiple content types including text, images, audio, and video.
- The platform provides a simple interface to experiment with prompts and integrate AI capabilities into products or workflows.

Why It Matters?

- Enables quick creation of high-quality content across different media formats.
- Helps developers and creators automate and scale creative processes.
- Offers an accessible environment to test multimodal AI capabilities in one place.

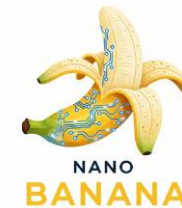
<https://aistudio.google.com/>

Google AI Studio

Content Type	Capabilities	Underlying Models / Tools
Text	Generate articles, scripts, chatbots, and summaries	Gemini text generation models
Image	Create visuals, illustrations, and designs from text prompts	Imagen models integrated into AI Studio
Audio	Produce voice narration, sound design, and speech synthesis	Gemini audio generation tools
Video	Generate short clips or animations from text and image inputs	Veo text-to-video models

Key Features

- Unified interface for all modalities.
- Adjustable parameters for creativity and style control.
- Integration-ready APIs for deploying generated content.
- Supports prototyping, testing, and refining generative outputs efficiently.



Limitations and Ethical Considerations

Hallucinations:

1. Definition: Generative AI models can confidently produce false, nonsensical, or made-up information.
2. Reason: They are trained to generate plausible sequences of words, not necessarily to be factual.
3. Mitigation: Always verify information generated by AI, especially for critical business decisions.

Bias:

1. Source: AI models learn from the data they are trained on. If the training data contains societal biases (e.g., gender, racial), the AI can perpetuate or amplify them.
2. Implication: AI-generated content might reflect these biases, leading to unfair or inappropriate outputs.
3. Mitigation: Be aware of potential biases and critically review AI outputs for fairness and inclusivity.

Data Privacy and Security:

1. Concern: When you input sensitive company data into public AI models, there's a risk of that data being used for training or exposed.
2. Best Practice: Avoid inputting confidential or proprietary information into general-purpose AI tools.

Copyright and Intellectual Property:

1. Issue: Who owns the content generated by AI? Can AI generate content that infringes on existing copyrights?
2. Current State: This is an evolving legal area. Be mindful of these complexities, especially for commercial use.

Test Drive LLMs

- <https://ai.google.dev/>

Create Gemini API Key at:

- <https://aistudio.google.com/app/api-keys>
- <https://ai.google.dev/gemini-api/docs/quickstart>